The development and use of machine translation systems and computer-based translation tools

John Hutchins
(University of East Anglia, Norwich, UK)
Email: WJHutchins@compuserve.com

Types of translation demand and systems

a) dissemination of information/documents (publishable quality)
           external publications
           internal reports
           operational manuals
           localization of software
   systems:
           'traditional' batch MT
              with post-editing and/or controlled input
           translation workstations
              with terminology databases and/or translation memories

b) assimilation of information (rough versions)
           internal purposes (information analysis)
           information service
   systems:
           'traditional' MT systems
              with no editing/revision
           PC-based software

c) interchange
           delayed: correspondence, electronic mail
           immediate: telephone
   systems:
           PC-based software
           speech translation

d) information access
           document/information retrieval
           information extraction
           summarisation
           database searching
           Web browsing
   systems:
           PC-based Web/HTML translations for Internet
           purpose-built

Historical background

1949 : Warren Weaver's memorandum
1954 : first public demonstration
1966: ALPAC report
1975 : Météo, Systran acquired by European Commission
1985 : Japanese commercial systems
1995: MT for e-mail and Web

Some systems

Systran, Logos, Atlas (Fujitsu)
Metal, Ariane, Eurotra, JICST (Mu), MMT
KANT (Carnegie-Mellon University), DLT, LUTE, Pivot
IBM (Candide), Pangloss
ATR, JANUS, C-Star, Verbmobil

From Perfectionism to Diversification

Fully automatic high quality translation (FAHQT)

Bar-Hillel (1960): demonstration of non-feasibility of FAHQT
                  recommendation of 'man-machine symbiosis'

neglect of:
      operational requirements
            (use of 'less than perfect' output)
      expertise of translators

ALPAC (Automatic Language Processing Advisory Committee), 1966
      Recommendations:
            No further funding of MT in the USA
            Development of translation aids

Since ALPAC, three strands of MT:
      Translation tools
            (translator workstations, translation memory)
      Operational MT systems (with human involvement)
            pre- and post-editing
            domain-specific
            controlled language
      MT research
            Linguistics-based approaches (transfer, interlingua)
            Knowledge-based systems
            Example-based MT
            Statistics-based MT
            Spoken Language Translation

Use of MT systems: types of use

Governmental and non-commercial use
      Readable ('draft') quality
      Rapid delivery
           Examples:
                 US Air Force (Systran); National Air Intelligence Center
                 European Commission

Technical documentation
      Good quality output
      Human revision ('post-editing')
      Large volumes
      Consistency in terminology
      Controlled language input
      Multiple target languages
      Integration in documentation processes
           Examples:
                 Logos: Lexi-Tech, Ericsson, Osram, Océ, SAP, Corel
                 Systran: Ford, General Motors, Aérospatiale, Berlitz, Xerox
                 Metal: SAP, Philips, Union Bank of Switzerland

Non-professional 'casual' use (monolingual)
      Lower quality
      Rapid delivery
           Examples:
                 Electronic mail, Web pages

Controlled language and customer-specific systems

Controlled language:
    Xerox Corporation: Multinational Customized English
    Caterpillar Corporation

Custom-built systems:
    Smart Corporation
    Cap Volmac
    LANT

Restricted language:
    LingTech A/S (PaTrans)
    Météo
    TITUS
    NHK

In-house:
    Pan American Health Organization (SPANAM, ENGSPAN)
    Kielikone (TranSmart)
    Hook and Hatton
    Japan Information Center for Science and Technology
    IBM Japan (SHALT)
    CSK (ARGO)

Special purpose:
    Diplomat (CMU)
    TCC Communications

Translator Workstation facilities

Multilingual word processing

Optical character recognition

Electronic receipt and transmission of texts

Terminology management

Access to local and remote dictionaries, terminology databanks and other information resources

Concordance software

Translation memory (and alignment software)

Bilingual text corpora

Automatic translation (MT system)

Documentation management

Integration with publishing software

Translation Workstations

TRADOS TWB

STAR Transit

IBM TranslationManager/2

Eurolang Optimizer (LANT)

Non-commercial:
    EURAMIS

Localization

SAP (METAL and Logos)

Corel

Lotus

Canon

LISA (Localisation Industry Standards Association)

SLIG (Software Localisation Group, Ireland)

OTELO

PC software

Power Translator (Globalink) *
Telegraph (Globalink) *
PC-Translator (Linguistic Products)
Transcend (Transparent Language)
LogoVista E J (Language Engineering Corporation)
Tsunami (Neocor Technologies) *
Typhoon (Neocor Technologies) *
TranSphere (AppTek) *
Winger (Winger A/S)
ProMT 98 (Project MT)
Reverso (Softissimo)
PARS (Lingvistica'98)
Lexica (EPI-USE Systems)
Systran Classic, Systran Personal, Systran Pro
Personal Translator PT (IBM/von Rheinbaben & Busch)
Langenscheidt T1 (GMS) *
iTranslator (Lernout & Hauspie)

* = now owned by Lernout & Hauspie

Japanese commercial systems

ATLAS (Fujitsu)
ASTRANSAC (Toshiba) [HONYAKU series]
HICATS (Hitachi) [Tamachi Honyaku; HICOM-MT]
PENSEE (Oki)
PIVOT (NEC) [Honyaku Adapter/Crossroad]
DUET (Sharp)
Transer (Nova) [PC-Transer]
TransLand (Brother)
Korya Eiwa (Catena)
Honyaku Kobo (Panasonic/Matsushita)
Translator-Mini
TransSupporter (Sanyo)

MT services on the Internet
(providing raw output or lightly post-edited)

Systran
Globalink
CompuServe
Logos
LANT
ATLAS
PIVOT
JICST

Integration of systems and tools

MT systems
translation memory
terminology management
authoring systems
publishing systems

examples:
EURAMIS (Service de Traduction, EU Commission)
CELEX, Eurodicautom, MultiTerm, Trados
SAP
Logos and METAL, SAPterm, STERM
OTELO interface
Océ Technologies
Logos, translation memory (XL8, Trados),
writing control
Xerox (Grenoble)
XTRAS
STAR Transit
workflow module
LANT
METAL, Eurolang Optimizer
LANTMaster (controlled language)

Staff Management Implications

New skills:
>Terminology database (entry validation)
>Translation memory (inclusion/exclusion, quality)
>Text alignment (appropriate or not)
>MT systems (when and whether to use)
>Documentation flow

Personnel:
>Project manager
>Language engineer
>Translation technician
>Linguistic assistants
>Translators

Some conclusions

MT is not translation as usually understood
MT is no more than symbol and structure manipulation
     (where input and output interpretable (by human reader) as
     text in two different languages)
MT is a computer-based 'tool' (facility) [like spelling checkers]
     For translators (but prefer TM tools)
     For cross-language communication
     For access information resources in unknown languages
To be used as required, successfully or not, appropriately or not
     (saving costs and/or time in certain circumstances)

Questions for MT/TM software:
     Not --- does it produce 'real' translations?
     Not --- does it produce good translations?
     But --- can I use the output?
     And --- will its output save me time or money?

HT and MT appropriateness: a summary

	HT	MT/TM
Dissemination		
Literary, legal	costly, interesting	not poss.
Scientific, technical (single)	costly (+ dict)	costly (+edit)*
Technical (multi-target)	costly, boring	costly(+edit/TM)*
Weather reports	v. repetitive	yes
Localisation	costly, repetitive	yes + edit/TM *
Advertisements	not done	yes (adequate)
Assimilation (gisting, rough)		
Scientific/technical, etc.	not done	yes (reasonable)*
Other non-literary (occasional)	not done	yes (poor)*
Information monitoring	costly	yes (adequate)*
Interchange		
Correspondence (business)	boring	yes, interactive
Correspondence (personal)	yes (boring)	yes (poor)
Electronic mail	not done	yes (poor)**
Web pages	not done?	yes (poor)**
Information access:		
Database searching	imposs.	yes (adequate)*
Summarising	rare	yes (adequate)*
Drafting documents (authoring)	not done	yes (accepted)*
Television captions/subtitles	no	yes (adequate)*
Spoken language		
Informal conversation	yes	no
Telephone enquiries	not usual	[in principle]
Formal interpreting	yes	no

* = growing demand
** = rapidly growing demand